BIOSTATISTICS AT THE NIH

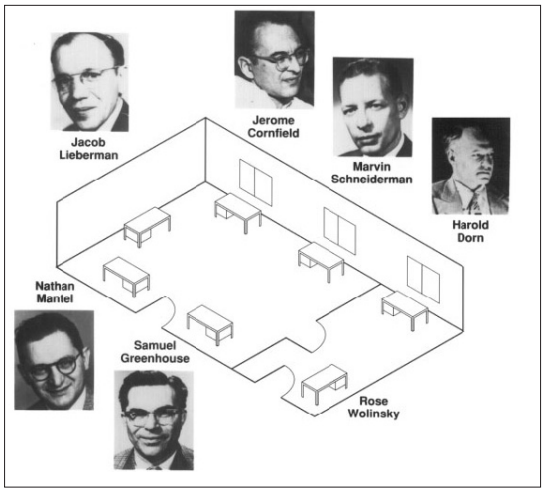
BY SEJAL PATEL, OFFICE OF NIH HISTORY

Biostatistics lies at the heart of just about every health investigation that takes place today. Investigators calculate sample sizes, ponder over confidence intervals, and pray for small *P*-values. NIH study sections and research journals alike expect investigators to use biostatistical methods in grant applications and papers. Most investigators would not think of proposing a study without at the very least consulting a biostatistician.

What investigators may not know is that NIH has played a pivotal role in the evolution of biostatistics, the application of statistics to analyze biological data and improve study designs. It all started in the late 1940s when Harold Dorn, who headed the U.S. Public Health Service's Division of Statistical Methods, recruited a talented team of statisticians—Nathan Mantel, Jerome Cornfield, Jacob Lieberman, Samuel Greenhouse, and Marvin Schneiderman—to the newly created Office of Biometry on the NIH campus. (The office soon became the Biometry Branch at the National Cancer Institute.)

When they came to NIH none had advanced degrees in statistics; most, however, had training during the war as operations researchers and so were adept at applying statistics to problem solving. For example, Schneiderman, when he worked for the U.S. Air Force, had helped to improve the productivity and efficiency of airplane engines by calculating the optimal time for the engines to be reconditioned.

With their training in applied statistics, problem solving, and population sampling, these biometrists excelled at optimizing methods for studying chronic diseases. They came to be known as the "gang of five," a loud, enthusiastic, argumentative group whose spirited debates could sometimes be heard in the Building One cafeteria. "Our luncheons were usually quiet and sociable, as long as we discussed subjects other than statistics. But when we raised statistical topics, almost always there were loud shouting matches without regard to the comfort of those around us," Greenhouse wrote in 1997. "At times, the arguments would carry over when we returned to the office and Dr. Dorn would stick his head through the doorway and innocently inquire what was going on." (*Stat Sci* **12**:82–87, 1997)



COURTESY NIH OFFICE OF HISTORY

Clockwise from bottom: The "gang of five" biostatisticians Samuel Greenhouse, Nathan Mantel, Jacob Lieberman, Jerome Cornfield, and Marvin Schneiderman and their fearless leader, Harold Dorn. The diagram represents where they worked: in Temporary Building 6, which used the concept of rooms without walls.

But as they engaged in research on statistical theory and methods, the design of experiments, applied probability, and applied mathematics, they were shaping the field of contemporary health research. In time, they developed an expertise in "the way in which one elicits the specific information from the investigator needed to find an optimum research design: the question, the nature of the measurements, the intervening factors that ought to be controlled, and so on," Greenhouse wrote.

The number of biostatisticians grew at NIH, and in 1949 biometry offices were established in each institute (a moment known as "the methodological big bang"). In 1953, NIH began funding university programs to produce more and better-qualified analytical statisticians for medical and health research. By 1960, more than 400 people had received NIH funding to pursue biostatistics training at either the masters or the doctorate level.

NIH Director James Shannon said in a 1959 speech that because of the rise in chronic disease, medical research "must include not only the physician but the whole array of university sciences; not only the individual patient but whole population groups; and that an essential and effective partner in the future is the sophisticated, biologically oriented mathematician."

The "gang of five" and other NIH biostatisticians left a lasting scientific legacy and designed some of the better-known methodological techniques still used by health investigators today. For instance, Cornfield's work estimating comparative rates from clinical data on cancer was considered one of the breakthroughs in biostatistics; his adaptation of multiple logistic regression analysis to health investigation is now a mainstay in risk-factor analysis. And Mantel's 1959 article on the Mantel-Haenszel test—"Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease"—was one of the most cited statistical articles for almost 20 years (*JNCI* **22**:719–748, 1959).

Shannon's comments from 1959 are as true now as they were then: "Biometry is now a powerful [language] in medical research, and for those who speak it well, it can open doors to new knowledge of health and disease." ●